Journal of Modern Applied Statistical Methods

Volume 1 | Issue 1 Article 9

5-1-2002

An Adaptive Inference Strategy: The Case of Auditory Data

Bruno D. Zumbo
University of British Columbia, bruno.zumbo@ubc.ca

Recommended Citation
Zumbo, Bruno D. (2002) "An Adaptive Inference Strategy: The Case of Auditory Data," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 1 , Article 9.
DOI: 10.22237/jmasm/1020255000
Available at: http://digitalcommons.wayne.edu/jmasm/vol1/iss1/9

Journal Of Modern Applied Statistical Methods
Winter 2002, Vol. 1, No. 1, 61-68

1538 - 9472/02/$30.00

An Adaptive Inference Strategy: The Case of Auditory Data

Bruno D. Zumbo

Department of Statistics
University of British Columbia

By way of an example some of the basic features in the derivation and use of adaptive inferential methods are demonstrated. The focus of this paper is dyadic (coupled) data in auditory and perceptual research. We present: (a) why one should not use the conventional methods, (b) a derivation of an adaptive method, and (c) how the new adaptive method works with the example data. In the concluding remarks we draw attention to the work of Professor George Barnard who provided the adaptive inference strategy in the context of the Behrens-Fisher problem — testing the equality of means when one doesn't want to assume that the variances are equal.

Keywords: Robustness, Inference, Statistical assumptions

Introduction

There are many uses of the expression "adaptive methods" in statistics and data analysis but, to my knowledge, all of them seek statistical procedures:

(i) good for a broad class of possible underlying models, but which are not necessarily best for any one of them,
(ii) where important parameters in the statistical procedure are specified after the sample is drawn, rather than fixed by prior considerations before the sample is observed, and
(iii) that let the sample data lead us toward plausible solutions to statistical problems.

Such adaptive methods are frequently characterized as being robust, that is, exhibiting strength in the face of real data situations where we know that most statistical models will seldom fit exactly the real situations; hence it does not seem productive to try to get the last ounce of mathematical efficiency out of some assumed situation. In my opinion, although he focused on estimation, the paper by Hogg (1974) is one of the clearest expositions of the basic tenants of adaptive methods.

The purpose of this article is to describe adaptive methods, in the context of an example, demonstrating both the derivation and application of adaptive methods. Unlike Hogg (1974), the focus of the present paper is adaptive inference. The example discussed herein is of the commonly found scenario of testing the equality of means for two independent groups. In the example, we concern ourselves with within-group correlation, wherein the conventional methods of inference fix this within-group correlation, by prior considerations, to zero — i.e., independent observations within groups. This example treats the problem of pairwise within-groups correlations; that is, coupled data.

Bruno D. Zumbo is Professor of Measurement, Evaluation, and Research Methodology as well as Associate Member of the Department of Statistics at the University of British Columbia. His areas of statistical research have included performance characteristics of standard procedures (both parametric and nonparametric) under non-standard conditions, as well as multivariate analysis and the statistical theories of measurement.

Coupled Data

Coupled data arise in the various fields of the social, behavioral, and health sciences. For example, relationship researchers regularly gather data from both members of the dyad (Kenny, 1995). The pairs can be heterosexual or homosexual couples, co-workers, family members or friendship pairs, to name a few examples. In perceptual research it is not unusual for researchers to report the number of organs (e.g., ears, eyes) tested, rather than the number of subjects. This latter situation, perceptual research, will be the focus of the present example.

In all of these cases, subjects or dyads are contributing two scores to the data pool. It can be reasonably argued that these two scores are not independent (i.e., uncorrelated) of one another. Data arising from such research should be referred to as coupled since each subject contributes a couplet of scores, and the correlation between these scores should be referred to as the intracouple correlation (Zumbo, 1996). This issue of coupled scores applies to audition, vision, and hemispheric laterality research, and any situation in which two lateral measures are made on one subject. Therefore, a defining characteristic of coupled data is that there are twice as many scores as there are subjects or dyads (i.e., there are n scores and $n/2$ subjects or dyads). Because the commonly used statistical inferential methods (not descriptive methods) assume that the n scores are independent, a potential problem may arise when a researcher bases their statistical analyses on the *n* scores ignoring that they arise from *n*/2 subjects or dyads. How, then, is one to perform inferential tests on data that are, potentially, highly interdependent — i.e., coupled data?

Before continuing with these new methods of analysis, I should perhaps take a closer look at the data structure for coupled data and discuss why we even need these new methods.

Coupled Data Structure

Coupled data arise in situations in which the observations in a study are not independent random variables, but rather are pairwise related. The researcher, however, is not interested in the differential effects of the elements of the pair. Coren and Hakstian (1990) initially brought this statistical problem to our attention in the area of auditory research. The statistical problem discussed by Coren and his associates has also been noted in vision research (Ederer, 1973; Rosner, 1982) and could conceivably occur in laterality studies, twin studies, or any experimental or quasi-experimental settings in which the assumption of independence within groups is violated by paired or, as I will refer to them, coupled data. Please note that what is being discussed here is obviously related to the units of analysis problem in survey or educational research wherein one deals with structured populations of respondents (e.g., clusters in sampling or classrooms in educational research). The methods presented herein could be extended to the classroom situation wherein one has more than two elements that are linked.

An Example

To illustrate the issues consider the data from a two-group completely randomized design given in Table 1. The data are from a hypothetical experiment reported in Zumbo and Zimmerman (1991) depicting auditory research. That is, assume an auditory researcher is interested in investigating whether there is a difference in hearing loss between two groups. The data is displayed in Table 1.

It is important to note that the researchers are not interested in differences between the left and right ears but rather they gather data from both ears and they are interested in group differences. Therefore, the researcher has a total of 12 observations (i.e., 6 couplets or dyads) in group 1 and 12 observations (i.e., 6 couplets) in group 2. In Table 1, I have placed a box around a couplet, furthermore the top score within the box is the left ear. Traditionally, this design has been envisioned as a two-group completely randomized design and analyzed with a parametric statistical test (for example, in this case, the independent samples t-test with 22 degrees of freedom) treating the data arising from the two members of the dyad as if they were independent (see Coren and Hakstian, 1990, for examples).

Table 1. Coupled data example.

Group 1		Group 2	
Dyad #	**X**	**Dyad #**	**X**
1	15.6	7	12.6
1	15.9	7	12.4
2	13.7	8	13.7
2	13.9	8	14.2
3	15.1	9	15.3
3	15.5	9	14.5
4	14.7	10	13.4
4	15.2	10	12.3
5	16.2	11	14.3
5	15.7	11	14.7
6	13.7	12	14.2
6	14.0	12	13.8
n = 12 mean = 14.93 std. dev. = 0.91		n = 12 mean = 13.78 std. dev. = 0.95	

What is wrong with treating this data with methods that fix the correlation to zero a priori?

The problem in dealing with these coupled data in this way is that for parametric tests a violation of within group independence can invalidate the statistical test (Zumbo, 1996; Zumbo & Zimmerman, 1991). More precisely, it can be shown mathematically that for t-tests and ANOVA, a positive correlation within couples results in an inflation in Type I error rate while a negative correlation results in a reduction in Type I error rate. Therefore, if the data from the two ears are positively correlated the Type I error rate is inflated; however, if the data from the two ears are negatively correlated the Type I error rate is deflated.

More formally, a function can be derived showing how the Type I error rate is altered by coupled data. The appendix provides further technical detail. Denote α as the nominal Type I error rate of the t-test (usually .05), and ε as the actual Type I error rate if we were to conduct the t-test incorrectly ignoring the coupled data, $n = n_1 = n_2$ denotes the common sample size, and ρ the intracouple correlation. The function is then written

$$t_\varepsilon = t_\alpha \sqrt{\frac{n-(1+\rho)}{(n-1)(1+\rho)}}\,. \tag{1}$$

Three points are noteworthy from the above equation. First, the amount that the Type I error rate is altered is a function of both the magnitude of the intracouple correlation and the sample size. Second, for a fixed sample size when $\rho=0$ $t_\varepsilon = t_\alpha$, while as ρ approaches negative one in the limit t_ε becomes larger than t_α, and as ρ approaches positive one in the limit t_ε becomes smaller than t_α. For example, for a nondirectional hypothesis test with 18 degrees of freedom $t_\alpha=2.10$, if $\rho=0$ then as expected $t_\varepsilon=2.10$; while for $\rho=-0.99999$ $t_\varepsilon=683.3$, and for $\rho=0.99999$ $t_\varepsilon=1.44$. Generally, then, if $\rho=0$ then $\varepsilon=\alpha$, a negative ρ would result in $\varepsilon<\alpha$, while a positive ρ would result in $\varepsilon>\alpha$. Finally, given that the distribution of *t* scores and the distribution of *F* scores are related by $t^2=F$, these results generalize to the fully randomized design ANOVA where,

$$F_\varepsilon = F_\alpha\left(\frac{n-(1+\rho)}{(n-1)(1+\rho)}\right).$$

Figure 1 is a graphical depiction of the relationship between Type I error rate and the correlation between the two observations that comprise the coupled data, ρ, for sample sizes of 4, 6, 8, and 10 and values of ρ ranging from -.90 to .90. It should be noted that the Type I error rates reported in Figure 1 are the complement of the cumulative density function for the central t with v degrees of freedom for the resulting $t_{v:\alpha}$ from equation (A10) — see the Appendix for details. The upper half of Figure 1 deals with a nondirectional model while the lower half deals with a directional model. First, it should be noted that the horizontal line traces the nominal Type I error rate and the vertical line traces ρ equal to zero. The intersection of the horizontal and vertical lines is the Type I error rate for the i.i.d. case. Second, the general relationship is the same for directional and nondirectional hypotheses. That is, a positive correlation results in inflation in Type I error rate, whereas a negative correlation results in a decrease in Type I error rate. Finally, sample size appears to have very little impact except in the case of a correlation of 0.60 or larger wherein the smaller sample sizes result in a slightly more inflated Type I error rate (a difference of approximately .02 to .06). The minimal effect of sample size is demonstrated in Figure 2.

Thus, if one ignores the fact that one has dyadic or coupled data then there can be a serious inflation (or possible deflation if the correlation is negative) in the Type I error rate of the test. This implies that an alternative method of analysis is needed.

An Adaptive inferential method

An adaptive method for analyzing the example data can be found by re-deriving the independent samples t-test allowing for a parameter in the t-test formula that measures the magnitude of the intracouple correlation, rather than apriori fix the correlation to zero. The

Figure 1. Type I error rates of the Student's t-test as a function of the correlation among the elements of the couple.

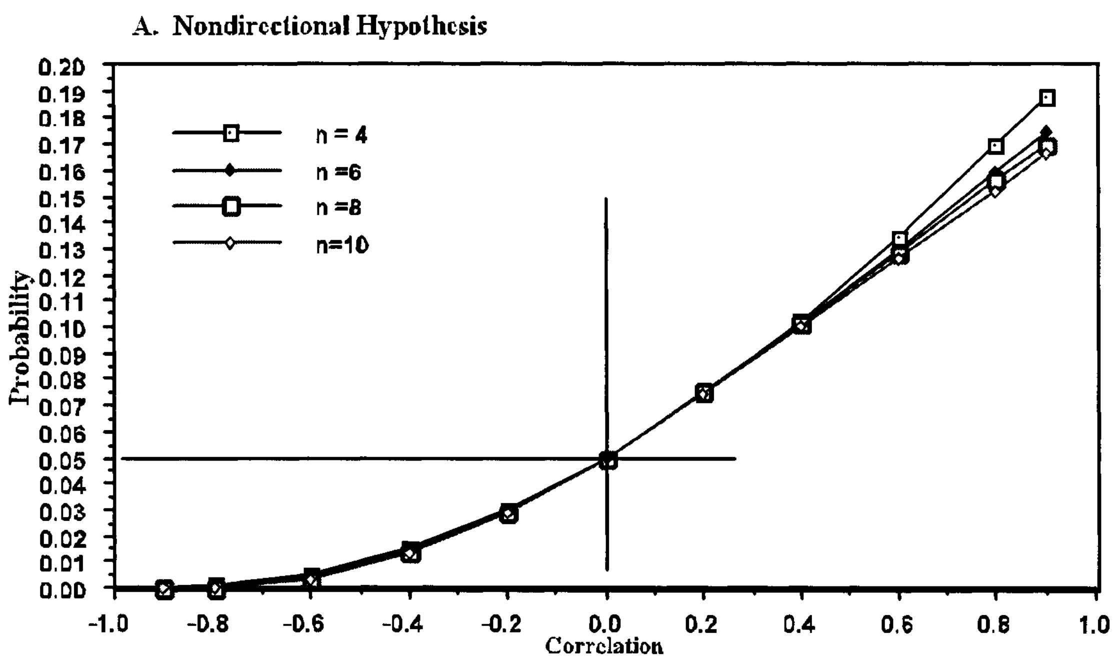


Figure 1 (Continued).

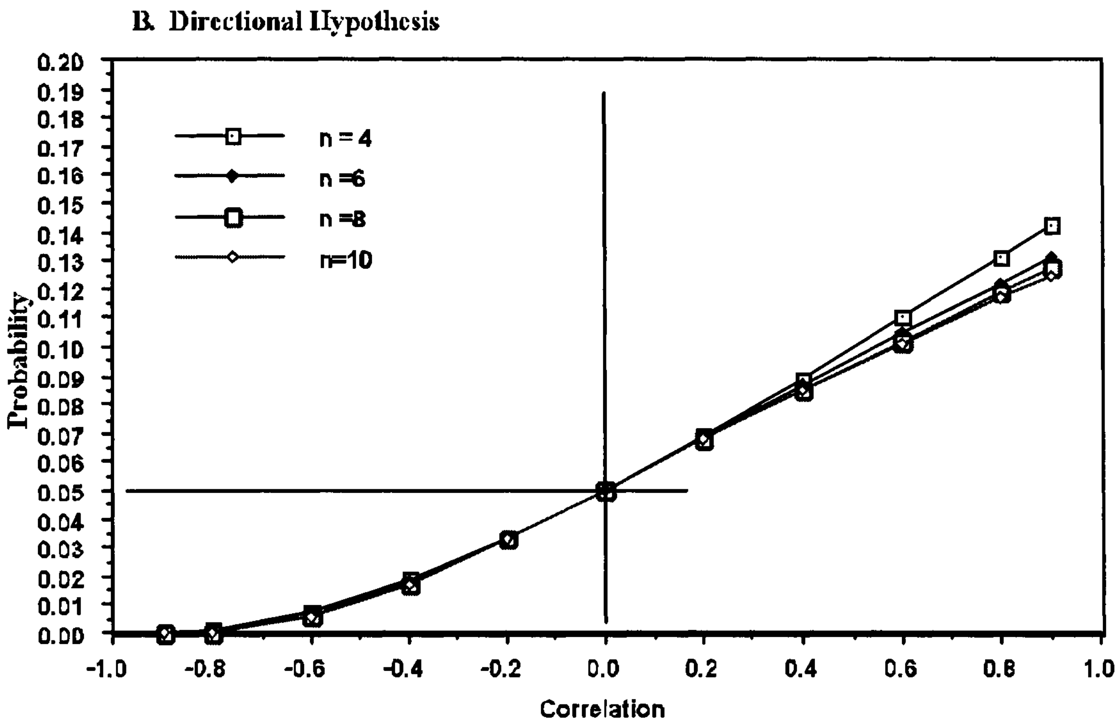
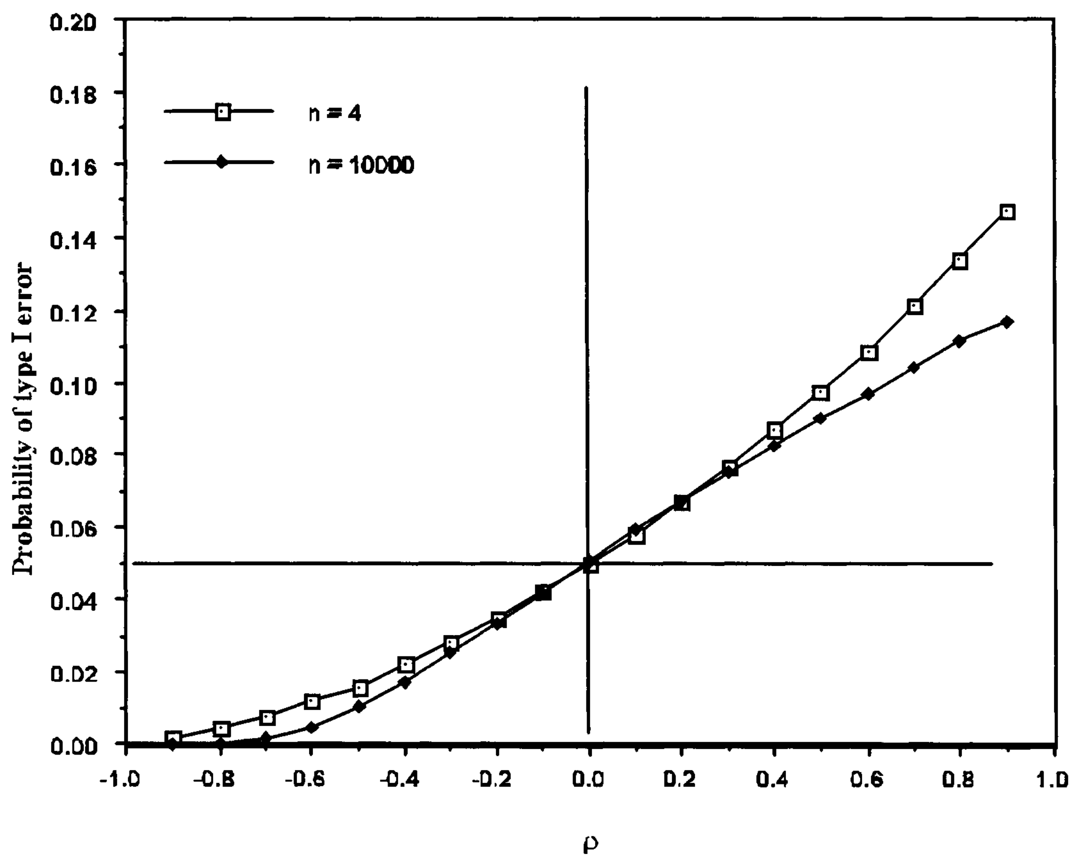


Figure 2. Type I error rates of the Student's t-test as a function of the correlation among the elements of the couple, n = 4 and 10,000 (Directional Hypothesis).



Appendix sketches such a derivation and leads to the replacement of the independent samples t-test by

$$t(\rho) = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{SS_{X_1} + SS_{X_2}}{n(n-1)} \times \frac{(n-1)(1+\rho)}{n-(1+\rho)}}}, \quad df = 2(n-1), \qquad (2)$$

wherein all of the symbols are described in the appendix and (2) applies for equal sample sizes and equal correlations for each group. Extending the strategy presented in the Appendix, one can derive the more general form allowing for unequal sample sizes and unequal correlations. The resulting more general t-test is

$$t(\rho_1\rho_2) = \frac{\overline{X}_1 - \overline{X}_2}{\frac{SS_{X_1}}{n_1} \times \left(\frac{(1+\rho_1)}{n_1-(1+\rho_1)}\right) + \frac{SS_{X_2}}{n_2} \times \left(\frac{(1+\rho_2)}{n_2-(1+\rho_2)}\right)},$$

$$df = n_1 + n_2 - 2 \qquad (3)$$

As an algebraic check, if the correlations for each group equal a common value, $\rho_1 = \rho_2 = \rho$, and the sample sizes for each group equal a common value, $n_1 = n_2 = n$, then after some algebraic rearranging (2) equals (3). Furthermore, if $\rho_1 = \rho_2 = 0$, then (3) simplifies to the standard unpooled version of Student's t-statistic for two independent samples.

For the purposes of our example we will use the t-test in equation (2). First we compute the common correlation between the left and right ears, r=.883, and then we compute a 90% interval for the correlation (.686, .959) using the so-called Fisher's r-to-z transformation and applying the formula $z_r \pm 1.645/\sqrt{N-3}$ where, in our case, N=12. Equation (2) can now be applied for the point and interval estimates of the correlation. Table 2 contains these three t-test results and the (incorrect) result when the correlation is equal to zero, t(0).

Clearly, it can be seen from Table 2 that there is no reason to suppose that the intracouple correlation is zero. Furthermore, it can be seen that the value of the test statistic is, as described earlier in this paper, sensitive to non-zero correlation. However, in presenting the results in the manner of Table 2, it can be assessed how sensitive the inference is to the assumption of zero correlation. If a nominal error rate of .05 is used, then the statistical decision is not effected by even a substantial non-zero correlation, whereas this would not be true for an error rate of 0.01. Finally, it is important to note that this sort of *sensitivity* analysis needs to be conducted for each data set you have because in some cases the statistical decision may be affected by even a slight non-zero correlation.

It should be noted that this data is hypothetical and was generated with a standardized difference between the population means of 1.50 (Zumbo & Zimmerman, 1991). That is, there is a substantial difference in the population means. (As a side note, a suggested method for analyzing this sort of data is to average across the two elements of the dyad and hence halving your sample size. This results in a statistically non-significant result, t(10)=2.13, *p*= 0.06.)

Conclusion

The purpose of this paper was to show how it might be more illuminating in day-to-day statistical applications to use an adaptive statistical strategy. For example, the adaptive t-test was computed for a plausible range of intracouple correlation values ranging from .686 to .959. This, I believe, sheds more light on the problem than simply averaging over the two elements of the couple, which is a commonly recommended strategy (see Coren & Hakstian, 1990) and resulted in a statistically non-significant finding that conceals the effect of intracouple correlation. The full range of correlations, including the point estimate, gives the analyst a sense of the dependence of the result on the

Table 2. The resulting t-test statistics at various values for ρ in Equation (2).

t(ρ)	t-value	degrees of freedom	p-value	magnitude of effect (point biserial correlation)
t(0)	3.03	22	.006	.295
t(.686)	2.26	22	.034	.189
t(.883)	2.12	22	.045	.170
t(.959)	2.07	22	.050	.163

intracouple correlation. A similar approach could be used to study the units of analysis (wherein students are clustered within classrooms) in educational research. One could apply the same sort of analytic strategy as used in the Appendix and derive a t-test parameterized by an intraclass correlation. In doing their data analysis one could then investigate plausible values of the intraclass correlation and see how these values alter the statistical conclusion.

It should be noted that the coupled data problem is not the only problem that has been dealt with as adaptive inference. In fact, the approach presented herein is a strategy developed by Barnard (1982, 1984). He gave a similar treatment to the Behrens-Fisher problem by presenting a t-test that has as a parameter the ratio of the sample variances (see, e.g., Sprott & Farewell, 1993).

Barnard showed that for the Behrens-Fisher case, the problem is to make inferences about the differences in means without fixing the ratio of the two variances, by prior considerations, to one. Barnard's method allows one to explore various values of the variance ratio (in fact, plausible values computed from the sample data, much like the intracouple correlation discussed above) and then one can see how constraining the value to one may, in fact, conceal the sensitivity that the t-test has to plausible values of the variance ratio. Although Barnard presented a method in the context of fiducial distributions, pivotals, robust pivotals, and pivotal likelihoods, the methods presented herein are an application of Barnard's analytic strategy of data-adaptive inference. In this data-adaptive inference, the data lead to sensible solutions.

References

Barnard, G. (1982). A new approach to the Behrens-Fisher problem. *Utilitas Mathematica*, *21*, 261-271.

Barnard, G. (1984). Comparing the means of two independent samples. *Applied Statistics*, *33*, 266-271.

Coren, S., & Hakstian, A. R. (1990). Methodological implications of interaural correlation: Count heads not ears. *Perception & Psychophysics*, *48*, 291-294.

Ederer, F. (1973). Shall we count number of eyes or number of subjects? *Archives of Ophthalmology*, *89*, 1-2.

Hogg, R. V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, *69*, 909-923.

Kenny, D. A. (1995). The effect of nonindependence on significance testing in dyadic research. *Personal Relationships*, *2*, 67-75.

Rosner, B. (1982). Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics*, *38*, 105-114.

Sprott, D. A., & Farewell, V. T. (1993). The difference between two normal means. *The American Statistician*, *47*, 126-128.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences (2nd Ed.)*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Zumbo, B. D. (1996). Randomization test for coupled data. *Perception & Psychophysics*, *58*, 471-478.

Zumbo, B. D., & Zimmerman, D. W. (1991). Further evidence for Coren and Hakstian's "Methodological implications of interaural correlation: Count heads not ears" and an alternative correction formula. *Perception and Psychophysics*, *50*, 297-301.

Appendix

In the commonly used model-based general linear model, a random sample of size n is a sequence of observations of independent identically distributed (i.i.d.) random variables, $X_1, X_2, \ldots, X_n$. Under this model

$$\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}, \qquad \text{(A1)}$$

$$E(S^2) = \frac{n-1}{n}\sigma_X^2, \text{ and} \qquad \text{(A2)}$$

$$\hat{\sigma}_X^2 = \frac{SS_X}{n-1}, \qquad \text{(A3)}$$

where (A1) is the variance of a sample mean, (A2) is the mean of a sample variance, and (A3) is an unbiased estimate of the population variance. Here, I use the notation

$$SS_X = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \text{ and}$$

σ_X^2 denotes the population variance of the sample observations.

The derivation of equations (A1), (A2), and (A3) is simplified by the fact that the covariance terms in the general equation for the variance of a sum, S_n,

$$\begin{aligned}\sigma^2(S_n) &= \sigma^2(X_1 + X_2 + \ldots + X_n)\\ &= \sigma^2(X_1) + \sigma^2(X_2) + \ldots + \sigma^2(X_n) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)\end{aligned}$$

are all zero; where $S_n = X_1 + X_2 + \ldots + X_n$.

This section derives expressions analogous to (A1), (A2), and (A3) that include nonzero covariance terms due to coupled data. If we let

$$\rho = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_X^2}, \quad i \neq j,$$

and ρ is the same for all i and j, then for coupled data it turns out that

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}[1+\rho], \tag{A4}$$

$$E(S^2) = \sigma_X^2 \frac{[n-(1+\rho)]}{n}, \text{ and} \tag{A5}$$

$$\hat{\sigma}_X^2 = \frac{SS_X}{n-(1+\rho)}, \tag{A6}$$

As in expressions (A1) to (A3), n in (A4) to (A6) denotes the number of observations — except in this case they are not i.i.d. but rather are coupled data. As an algebraic check, if $\rho = 0$, (A4), (A5), and (A6) reduce to (A1), (A2), and (A3), respectively.

One can now use (A4), (A5), and (A6) in lieu of their corresponding i.i.d. expressions to derive a Student's t-test for the balanced two-group completely randomized design assuming a common ρ for both groups. That is, one can place a two-sided confidence interval around $(\mu_1 - \mu_2)$ by using

$$(\overline{X}_1 - \overline{X}_2) \pm t_{v:\alpha/2}\sqrt{\frac{SS_{X_1} + SS_{X_2}}{n(n-1)} \times \frac{(n-1)(1+\rho)}{n-(1+\rho)}}, \; v = 2(n-1) \tag{A7}$$

where v denotes the usual degrees of freedom, and n denotes the common sample size. Equation (A7) is re-expressed as equation (2) in the main body of the text, a t-test of the two independent groups balanced design.

Interestingly, applying Cochran's theorem (1934; Searle, 1971, Sections 2.5 and 2.6; 1982, p. 356) regarding the distribution of quadratic forms to (A7), it can be shown that (A7) is not distributed as t and is therefore an approximate test. However, Zumbo and Zimmerman (1991) showed via Monte Carlo simulation that (A7) is an adequate approximation, maintaining its empirical Type I error rate very close to its nominal value. One can gain insight into how the approximation works by noting that the expected value of the variance, can be expressed as

$$E(s^2) = \left(\sigma_X^2 \times \frac{n-1}{n}\right) \times \left(1 - \frac{\rho}{n-1}\right),$$

and is clearly asymptotically unbiased.

Now, given (A7), I turn to the task of deriving a general expression indicating the severity of the alteration to the Type I error rate. Given that (A7) is an approximate test, the following results are not exact, but rather good approximations and should be indicative of the behavior of the Type I error rate.

Without loss of generality, let us consider the one-sided confidence interval computed for the population mean difference, $(\mu_1 - \mu_2)$. Given the i.i.d. assumption, the one-sided confidence interval for small samples is denoted by

$$\Pr\left[(\overline{X}_1 - \overline{X}_2) + t_{v:\varepsilon}\sqrt{\frac{SS_{X_1} + SS_{X_2}}{n(n-1)}} < (\mu_1 - \mu_2)\right] = \varepsilon \qquad \text{(A8)}$$

where $t_{v:\varepsilon}$ equals the 100(ε) percentile of the t distribution with $v = 2(n\text{-}1)$. Now, given coupled data (A8) can be rewritten as

$$\Pr\left[(\overline{X}_1 - \overline{X}_2) + t_{v:\alpha}\sqrt{\frac{SS_{X_1} + SS_{X_2}}{n(n-1)} \times \frac{(n-1)(1+\rho)}{n-(1+\rho)}} < (\mu_1 - \mu_2)\right] = \alpha \qquad \text{(A9)}$$

where $t_{v:\alpha}$ denotes the t value exceeded by probability α. It should be noted that α is the nominal level of the test and ε is the actual level achieved due to not accounting for the covariance due to coupled data.

Finally, setting equation (A8) equal to (A9) results in,

$$t_{v:\alpha} = t_{v:\varepsilon}\sqrt{\frac{n-(1+\rho)}{(n-1)(1+\rho)}}. \qquad \text{(A10)}$$

If the n observations are i.i.d., then $\rho = 0$ and $\alpha = \varepsilon$. Therefore, if $\rho \neq 0$, then α can be quite different from ε. As noted above, (A10) can be used with directional or nondirectional hypotheses. Equation (A10) is listed as equation (1) in the main text of this paper.